Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1648 (2003) 1–11
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Crystallographic structure at 1.6-Å resolution of the human adenovirus proteinase in a covalent complex with its 11-amino-acid peptide cofactor: insights on a new fold

William J. McGrath[a], Jianzhong Ding[b], Aashish Didwania[a], Robert M. Sweet[a], Walter F. Mangel[a,*]

[a] *Biology Department, Brookhaven National Laboratory, 50 Bell Avenue, Upton, NY 11973-5000, USA*
[b] *Physics Department, State University of New York at Stony Brook, Stony Brook, NY 11794, USA*

Received 19 February 2002; received in revised form 21 October 2002; accepted 13 January 2003

Abstract

The crystal structure of the human adenovirus proteinase (AVP), a cysteine proteinase covalently bound to its 11-amino-acid peptide cofactor pVIc, has been solved to 1.6-Å resolution with a crystallographic *R*-factor of 0.136, R_{free} = 0.179. The fold of AVP–pVIc is new and the structural basis for it is described in detail. The polypeptide chain of AVP folds into two domains. One domain contains a five-strand β-sheet with two peripheral α-helices; this region represents the hydrophobic core of the protein. A second domain contains the N terminus, several C-terminal α-helices, and a small peripheral anti-parallel β-sheet. The domains interact through an extended polar interface. pVIc spans the two domains like a strap, its C-terminal portion forming a sixth strand on the β-sheet. The active site is in a long, deep groove located between the two domains. Portions are structurally similar to the active site of the prototypical cysteine proteinase papain, especially some of the Cα backbone atoms (r.m.s. deviation of 0.354 Å for 12 Cα atoms). The active-site nucleophile of AVP, the conserved Cys^{122}, was shown to have a pK_a of 4.5, close to the pK_a of 3.0 for the nucleophile of papain, suggesting that a similar ion pair arrangement with His^{54} may be present in AVP–pVIc. The interactions between AVP and pVIc include 24 non-β-strand hydrogen bonds, six β-strand hydrogen bonds and one covalent bond. Of the 204 amino acid residues in AVP, 33 are conserved among the many serotypes of adenovirus, and these aid in forming the active site groove, are involved in substrate specificity or interact between secondary structure elements.

Keywords: Viral proteinase; Peptide cofactor; Cysteine proteinase; X-ray crystallographic refinement; Active-site nucleophile

1. Introduction

The human adenovirus serotype 2 proteinase (AVP), like many virus-coded proteinases, is required for the synthesis of infectious virus [1]. Late in infection, young virions are formed in which 6 of the 12 major proteins are precursor proteins. In the young virion, the approximately 70 molecules of AVP become activated [2] and cleave multiple copies of these six virion precursor proteins more than 3200 times, thereby rendering the virus particle infectious [3]. A temperature-sensitive mutant of adenovirus was shown to lack proteinase activity at the nonpermissive temperature [1]; the mutation mapped to the L3 23K gene [4]. The L3 23K gene was cloned and expressed in *E. coli* [5,6] and baculovirus-infected cells [7], and the resultant 204-amino-acid protein, AVP, was purified [5,7].

Purified recombinant AVP is inactive, although the proteinase in disrupted virions is active [5]; this eventually prompted a search for a cofactor. Two were discovered. One cofactor is pVIc, the 11-amino-acid peptide whose sequence is GVQSLKRRRCF [5,7]. It originates from the C terminus of the virion precursor protein pVI. The four amino acid residues preceding pVIc in pVI constitute an AVP consensus cleavage sequence. Thus, AVP can cleave out its own cofactor. The other cofactor is the viral DNA [5]. Binding of the cofactors to AVP alters the interaction of the enzyme with the fluorogenic substrate (Leu–Arg–Gly–Gly–$NH)_2$–Rhodamine. In the absence of any cofactor, the K_m is 94.8 μM and the k_{cat} is 0.002 s^{-1} [8]. In the presence of a saturating amount of Ad2 DNA, the K_m decreases 10-fold

* Corresponding author. Tel.: +1-631-344-3373; fax: +1-631-344-3407.
E-mail address: Mangel@BNL.Gov (W.F. Mangel).

1570-9639/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S1570-9639(03)00024-4

and the k_{cat} increases 11-fold; in the presence of a saturating amount of pVIc, the K_m decreases 10-fold and the k_{cat} increases 118-fold. With both cofactors present, the k_{cat}/K_m ratio increases 34,000-fold compared to that for AVP alone.

AVP represents a new class of proteinase. The sequence of the gene for AVP was not related to any gene sequence in the databases [9–11], until recently [12]. Inhibitor profiles of enzyme activity are inconclusive with regard to the type of proteinase [13]. The crystal structure of the AVP–pVIc complex at 2.6 Å resolution was solved with single-isomorphous replacement supplemented with anomalous scattering [14]. The fold of the protein had not been seen before. Comparison of the structure of AVP–pVIc with the structures in the Protein Data Bank [15] revealed no equivalent structure, suggesting that AVP represented a new family of protein molecules. However, it was shown to represent an example of either subtly divergent or powerfully convergent evolution, because the active site contains a Cys–His–Glu triplet and oxyanion hole in an arrangement similar to that in the archetypical cysteine proteinase papain [14]. Thus, AVP represents a new, fifth group of enzymes that contains catalytic triads. pVIc, which extends a β-sheet in the main chain, was found to be quite distant from the active site. There are four regions of high positive charge density that are potential DNA binding sites. Modeling of a substrate in the active site as an extended β-strand revealed the structural origin of substrate specificity.

Adenoviruses are the cause of a number of acute infections, especially of the respiratory and gastrointestinal tracts, and of conjunctivitis. Because of the large number of human adenovirus serotypes, more than 50, it may be difficult to obtain a universal adenovirus vaccine. Hence, there is a need for antiviral agents against adenovirus. AVP, like other virus-coded proteinases, is a highly specific enzyme, and this makes it attractive as a target for antiviral agents [16,17]. Discovery of antiviral agents by structure-based drug design requires high-resolution crystal structures. Elucidation of the structural basis for the regulation of enzyme activity also requires high-resolution structures. Here, we report the details of the refinement of the crystal structure of the AVP–pVIc complex at high resolution, 1.6 Å, and describe the major features of the enzyme from the points of view of drug design and regulation of enzyme activity.

2. Materials and methods

2.1. Protein concentrations

Recombinant AVP was purified as described previously [18]. The concentration of AVP was determined using a molar absorbance coefficient at 280 nm of 26,510, calculated according to the method of Gill and von Hippel [19]. The concentration of pVIc (GVQSLKRRRCF) was determined by titration of its cysteine residue with Ellman's reagent and confirmed by quantitative amino acid analysis. Ellman's assays were performed in a volume of 0.516 ml in 50 mM Tris (pH 8.0) containing 350 μM DTNB and pVIc. Reactions were incubated at 25 °C for 10 min before the absorbance at 412 nm was measured. The concentration of thiolate anion was calculated by subtracting the absorbance in an assay lacking pVIc from that measured in its presence, using a molar extinction coefficient for thionitrobenzoate at 412 nm of 14,150 [20]. pVIc was purchased from Research Genetics, Huntsville, AL.

2.2. AVP–pVIc complex formation

AVP–pVIc complexes for activity assays were formed by incubation of either 25 μM AVP with 65 μM dimeric pVIc or 100 μM AVP with 150 μM monomeric pVIc, at 4 °C for 0.5 to 16 h. All buffers were saturated with nitrogen.

2.3. Proteinase activity assays

The fluorogenic substrate $(Leu–Arg–Gly–Gly–NH)_2$–Rhodamine was synthesized as described [5]. Assays were carried out using either an ISS PC-1 spectrofluorometer (Urbana, IL) equipped with a thermostatted sample chamber or a Photon Technologies International (PTI) QuantaMaster spectrofluorometer (Monmouth Junction, NJ) equipped with a thermostatted sample chamber. For both instruments, the excitation wavelength was 492 nm and the emission wavelength was 523 nm. For the ISS spectrofluorometer, the band pass was 8 nm, and for the PTI spectrofluorometer, the band pass was 4 nm.

2.4. Carboxamidomethylation of AVP, AVP–pVIc and papain as a function of pH

Carboxamidomethylation reactions were performed in a total volume of 0.02 ml in 0.1 M pH buffer. To achieve a pH between pH 3.0 and 4.0, the buffer was sodium formate; between pH 4.5 and 5.5, the buffer was sodium acetate; between pH 6.0 and 6.5, the buffer was MES; between pH 7.0 and 7.5, the buffer was HEPES; for pH 8.0, the buffer was Tris. All buffers were saturated with nitrogen.

AVP, 12.5 μM, was fully activated by incubation in 5 mM sodium phosphate (pH 8.0), 25 mM NaCl, 0.1 mM EDTA and 1 mM DTT at 25 °C for 1 h. Then, AVP and AVP–pVIc were incubated at a concentration of 3 μM in 0.1 M pH buffer for 2 min at 25 °C in the absence or presence of 2 mM iodoacetamide.

Papain, 165 μM, was activated by incubation in 25 mM sodium acetate (pH 5.0) and 10 mM DTT for 1 h at 25 °C. The papain was then diluted to 20 μM in 10 mM TAPS (pH 8.5) and 1 mM EDTA. For the carboxamidomethylation experiment, papain, at 2 μM, was incubated at 25 °C for 2 min in 0.1 M pH buffer in the absence or presence of 2 mM iodoacetamide.

2.5. Iodoacetamide inhibition assays

Activity assays for AVP were performed in a volume of 1 ml. AVP or AVP–pVIc from the carboxamidomethylation reactions was diluted to 15 nM into cuvettes containing 10 mM Tris, pH 8.0, 0.005% NP-40, 0.5 mM DTT, 286 ng/ml T7 DNA and 500 nM monomeric pVIc and incubated at 25 °C for 3 min. Then (Leu–Arg–Gly–Gly–NH)$_2$–Rhodamine was added to 10 μM and the increase in fluorescence monitored as a function of time.

Activity assays for papain were performed in a volume of 1 ml. Papain from the carboxamidomethylation reactions was diluted to 10 nM into cuvettes containing 25 mM sodium acetate (pH 5.0) and 30 μM (Cbz–Leu–Arg–NH)$_2$–Rhodamine and the increase in fluorescence monitored as a function of time.

2.6. Crystallization and data collection

The AVP–pVIc complex was formed and crystallized as described previously [21]. Prior to data collection, crystals were incubated in 0.11 M sodium HEPES, pH 7.5, and 1.54 M sodium acetate; glycerol was added in 4% (v/v) increments at 5-min intervals until its concentration reached 20%. The crystals were placed in cryoloops and flash-frozen in a 100 K nitrogen stream (Oxford Cryosystems, Oxford, UK).

X-ray diffraction data were collected at beamline X12-C of the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY). A wavelength of 1.15 Å was used. Data were collected on a MAR Research 300-mm diameter imaging-plate detector mounted on the 2θ arm of a FAST (Nonius) diffractometer. Frames were integrated, scaled, and reduced using the DENZO/SCALEPACK suite of programs [22]. Statistics on the data collection are in Table 1. Crystals formed in the hexagonal space group $P6_1$ with one molecule in the asymmetric unit. The unit cell dimensions of crystals held at approximately 100 K were $a=b=111.8$ Å, $c=50.9$ Å, $\alpha=\beta=60°$, $\gamma=120°$. For comparison, the unit cell dimensions in the original 2.6-Å resolution structure, measured at approximately 280 K, were $a=b=114.2$ Å, $c=50.1$ Å, $\alpha=\beta=90°$, $\gamma=120°$ [14].

Table 1
Crystallographic data and refinement statistics

Data collection	
Number of measured reflections	166,648
Number of unique reflections	44,447
Resolution (Å)	30–1.6 (1.66–1.60)
Completeness (%)	92.5 (77.1)
R-merge (%)	4.5 (16.6)
Percent of reflections with $I>2\sigma$ (I)	95.9 (88.2)
Average $I/2\sigma$ (I)	15.5 (7.6)
Mosaicity	0.65
Refinement	
Resolution range (Å)	30–1.6 (1.66–1.60)
Reflections used for *R*-factor	40,067
Reflections used for free *R*-factor	4,380
R-factor	0.136
Free *R*-factor	0.179
r.m.s. (bond lengths) (Å)	0.019
r.m.s. (bond angles) (°)	2.1
Protein residues	
AVP	203
pVIc	11
Acetate molecules	1
Water molecules	238
Isotropic temperature factors (Å^2)	
Protein atoms	14.09
Acetate molecule	13.55
Water molecules	32.30
Ramachandran Plot	
Most favored (%)	90.3
Additionally allowed (%)	9.7

Values in parentheses correspond to the highest resolution shell.

2.7. Structure determination and refinement

The 1.6-Å structure was solved by molecular replacement, because the unit cell dimensions in the crystals were different from those in the crystals used to obtain the original 2.6-Å resolution model [14]. Because of the high degree of isomorphism with the lower resolution crystal, we initially performed a translation search against the entire 2.6-Å resolution model. From this result, we then initiated three-dimensional refinement. The first step was rigid body refinement against the 2.6-Å resolution model. The refinement course was started with X-PLOR [23] using a limited resolution shell, first at 2.6 Å and then stepwise extended to higher resolution limits, to 2.2, to 1.9, and finally to 1.6 Å. Solvent was added. The final X-PLOR model for the AVP–pVIc complex contained 203 of the 204 AVP amino acids, 11 pVIc amino acids and 116 solvent molecules. It had an R_{cryst} of 0.228 and an R_{free} of 0.245, for data observed from 30 to 1.6 Å.

Further refinement of the model was performed using SHELXL [24] with all cycles of refinement restricted to the 6.0 to 1.6 Å resolution range. Ten conjugate-gradient cycles were performed each time. The mean distance from ideal values was 0.008 Å for X-PLOR and 0.019 Å for SHELXL. During the SHELXL refinements, several manual adjustments to the model were performed, such as remodeling the first seven amino-terminal residues, assignment of multiple conformers to three residues, introduction of an acetate molecule, and several rotamer corrections. After six cycles of SHELXL were performed, including the addition of solvent (whose occupancies were not refined), the low resolution range was extended to 30 Å and the resulting R_{cryst} was 0.136 and the R_{free} 0.179. The overall assessment of model quality was performed using PROCHECK [25].

2.8. Protein data bank accession number

Crystallographic coordinates for AVP–pVIc will be deposited with the RCSB Protein Data Bank as entry 1NLN.

3. Results and discussion

3.1. Quality of the model

The quality of the AVP–pVIc model was assessed as follows: The crystallographic *R*-factor from the last cycle of SHELXL [24] was 0.136 using 40,067 reflections from 30 to 1.6 Å; the R_{free} based on 4380 reflections was 0.179. The r.m.s. deviations from ideality for bond lengths and angles were 0.019 and 2.1°, respectively. A complete listing of the statistics of the refinement is presented in Table 1.

Comparisons of the $2|F_o| - |F_c|$ maps from a representative region of the 2.6- and 1.6-Å structures are shown (Fig. 1) to indicate the difference in quality of the maps. Inspection of the $2|F_o| - |F_c|$ maps of the N-terminal region of the complex resulted in remodeling of the first seven residues. The amino terminal methionine could not be found. Beginning at the glycine at position 2, which was clearly evident, the main chain was well defined through the rest of the molecule. This allowed the rebuilding of the remaining six residues of the N terminus. Unlike in the 2.6-Å structure, where the N-terminal Met^1–Gly^2 extended out from the surface of the structure, the amino terminal region in the 1.6-Å structure, beginning with Gly^2, clearly formed one of the walls of the active site groove and extended from near the nucleophile, Cys^{122}, to one end of the groove. The groove is closed after the P_4 binding site by the ionic bonds between the remodeled Glu^5 and Arg^{48}. The model becomes consistent with the 2.6-Å structure after the turn at Glu^7. Three side-chains were found with multiple conformations, serine residues 113 and 176, and lysine 81. Analysis of the main-chain dihedral angles with PROCHECK showed 90.3% of the non-proline, non-glycine residues fell in the most favored regions of the Ramachandran plot [25]. The remaining 9.7% of the non-proline, non-glycine residues were found in additional allowed regions. The average temperature factor for main-chain atoms was 12 $Å^2$; 14 $Å^2$ for side-chain atoms. The average temperature factor for water molecules was 32 $Å^2$. The overall temperature factor for the AVP–pVIc complex was 14.1 $Å^2$. The region with the highest temperature factors was located in the turn at residues 48–51. The estimate of coordinate error, based on the R_{free} [26], was 0.15 Å.

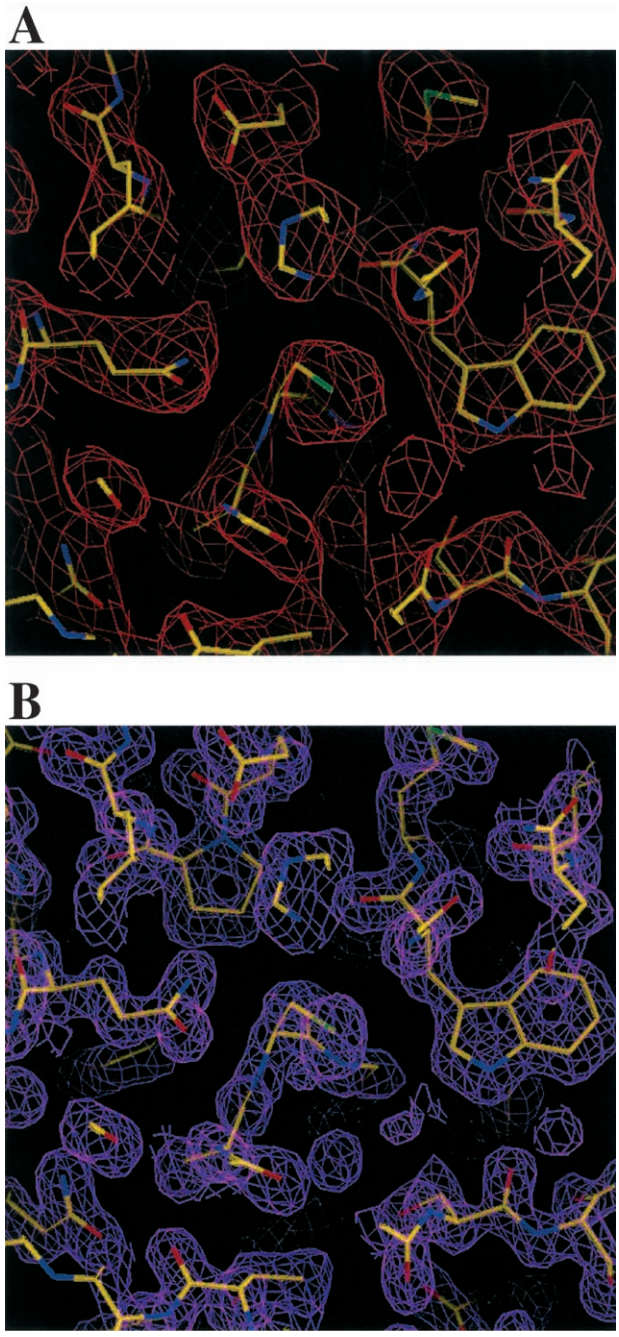

Fig. 1. Electron density of (A) 2.6 Å and (B) 1.6 Å $2|F_o| - |F_c|$ maps of the AVP–pVIc complex in the region near residue Trp^{55}, contoured at 1σ.

3.2. Analysis of the structure

The AVP–pVIc complex is ovoid with molecular dimensions of approximately 41 × 44 × 55 Å. The complex contains eight β-strands, including a portion of pVIc, and seven α-helices (Fig. 2A). There is a central β-sheet, containing five β-strands from AVP with a sixth strand from pVIc (residues 5′ to 9′), surrounded by two peripheral α-helices. The complex is folded into two domains (Fig. 2B). One domain contains the β-sheet, essentially residues 20 through 110 of AVP, the C-terminal seven residues of pVIc and two peripheral α-helices. The other domain consists mostly of α-helices, one from the N terminus and the remainder from the C-terminal portion of AVP. The domains interact through an extended polar interface. Residues 17 to 21 and 111 to 114 of AVP appear to act as two bridges that connect the domains. pVIc appears to act like a ‘strap’ that may assist in positioning the two domains. The first four residues of pVIc interact with

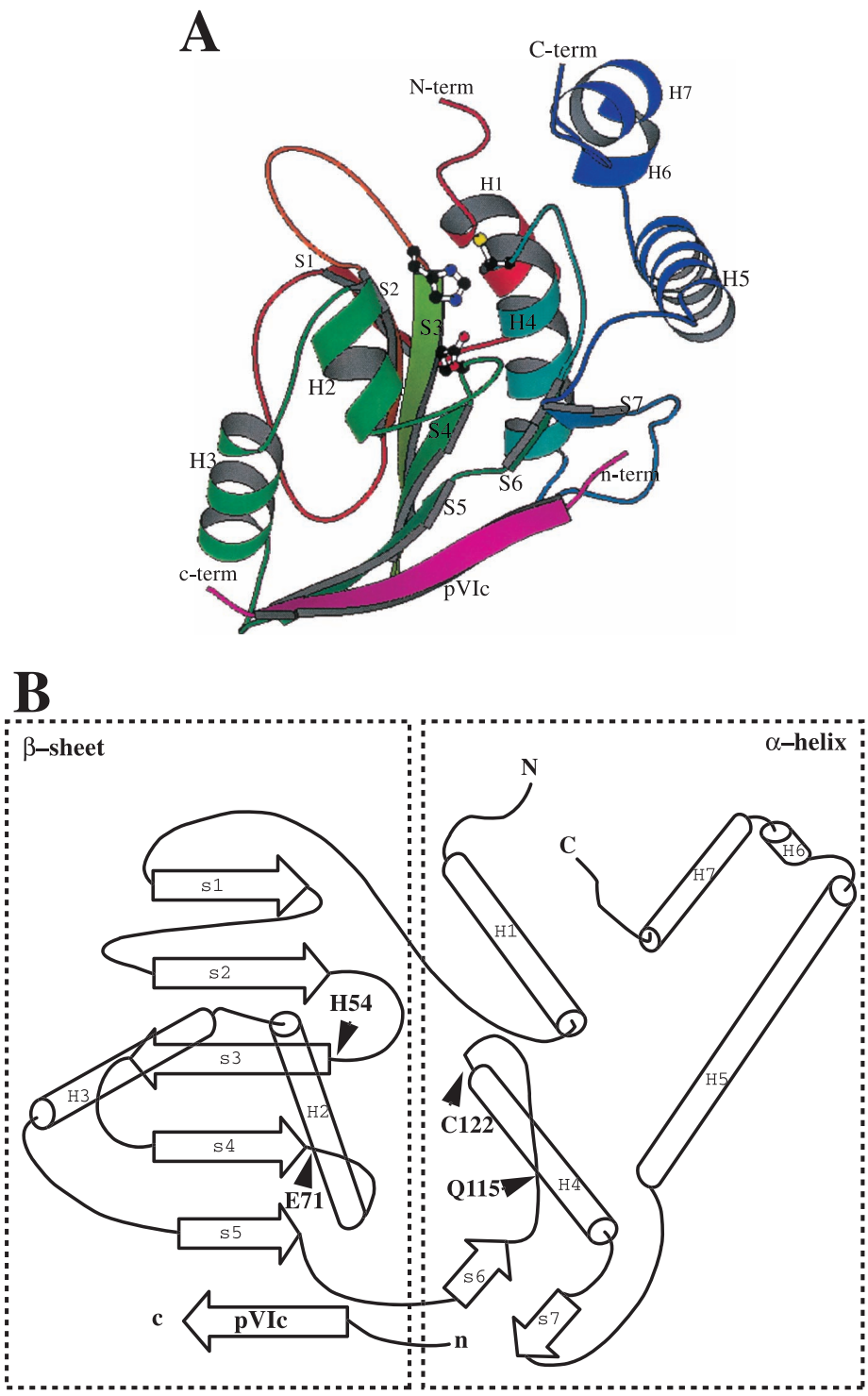

Fig. 2. The secondary structure of the AVP–pVIc complex. (A) The path of the polypeptide chain of AVP is color-coded according to the visible spectrum, red at the N terminus to violet at the C terminus. The α-helices are labeled H1 through H7, β-strands S1 through S7 from the N to C terminus. The pVIc peptide, colored magenta, is at the forefront of the β-strands in the figure. (B) A topological diagram of the secondary structure of the AVP–pVIc complex. Helices are shown as rods, β-strands as arrows. The length of the rod or arrow corresponds to the number of residues found in the structural element. The positions of the active-site residues are labeled. The dashed boxes indicate the two domains. The pVIc peptide and two loops from AVP are the only elements that connect the domains.

Fig. 3. Superimposition of the active site region of AVP–pVIc with papain. The backbone of papain (PDB Code 1PPN) in the vicinity of its active site is yellow and is superimposed onto the backbone of AVP–pVIc which is in green. The side chains of the active site residues of papain (magenta) overlay those of AVP–pVIc (blue).

components of the α-helical domain; the remaining seven residues are part of the β-sheet domain. The penultimate cysteine residue of pVIc forms a disulfide bond with Cys^{104} of AVP. There is a deep, curved groove nearly bisecting the molecule that contains the active site.

3.3. The active site groove

Although the fold of AVP is different from the fold of papain, in certain places there are similarities in architecture between the two proteins. In parts of the two active sites, there are similarities in secondary structure [14] (Fig. 3). Within the surface groove that nearly bisects the molecule are the central helix (H4) and several β-strands (S3 to S5). The active-site groove of the cysteine proteinase papain also contains a central helix and several β-strands. When these secondary structure elements in AVP–pVIc are aligned with the same secondary structure elements in papain, the amino acid residues involved in catalysis by papain (Cys^{25}, His^{159}, Asn^{175}, and Gln^{19}) align with similar residues in AVP–pVIc (Cys^{122}, His^{54}, Glu^{71} and Gln^{115}) (Fig. 3). This juxtaposition of elements involved in catalysis strongly suggests that AVP employs the same catalytic mechanism as papain.

The region most similar in structure between papain and AVP–pVIc is the active site region. In papain this is comprised of the central helix that contains the nucleophilic Cys^{25} in the active site groove. In both papain and AVP–pVIc, this central helix begins with the cysteine residue and extends through the core of the structures as part of the domain interface. Overlaying a portion of the central helix region of the papain structure (PDB code 1PPN) with the structure of the AVP–pVIc complex, using the lsq option of the program O [27], revealed that additional residues of papain juxtapose with residues in AVP–pVIc. For amino acid residues 23–34 of the central helix of papain, the r.m.s. deviation of the fit for the 12 Cα atoms to amino acid residues 120–131 of the AVP–pVIc complex was 0.354 Å. With the exception of the cysteine residues at the N termini of the helices, no other side chain residue superimposed. Extending the fit resulted in a total of 64 Cα atoms aligned with a resultant r.m.s. deviation of 2.104 Å, Table 2. Inspection of the superimposed structures revealed nearly identical positioning of only four side chains (Fig. 3). The side chains of Cys^{122} and His^{54} of the AVP–pVIc complex superimpose with Cys^{25} and His^{159} of papain with r.m.s. deviations of 0.79 Å for the eight atoms. For Glu^{71} and Gln^{115}, the critical hydrogen bonding side-chain atoms superimpose. The extent of the superposition of these residues was such that the distances between the catalytic residues of papain are nearly identical to those found in the AVP–pVIc structure, Table 3.

The composition of the active site groove of AVP–pVIc (Fig. 4) is as follows: Residues 51 to 55 form a portion of the wall and part of the floor of the active site groove. One of these amino acid residues, His^{54}, is the general base involved in catalysis. The groove is effectively closed at this end by the ion pair between Glu^{5} and Arg^{48} just beyond the side chain of the conserved Trp^{55}. Our model-building studies (see below) suggest that a consequence of closing the groove is the formation of a small pocket that

Table 2
Amino acid residues of papain (PDB access 1PPN) whose Cα atoms can be aligned with Cα atoms of AVP–pVIc

Papain	AVP–pVIc
3–6	104–107
10–14	138–142
21–40[a]	118–137
64–67	2–5
67–75	7–15
130–135	41–46
157–163	52–58
173–177	69–73
189–192	87–90

The initial fit was with the first 12 Cα atoms of H4 of AVP–pVIc onto papain. Then, the fit was improved, resulting in a total of 64 Cα atoms with an overall r.m.s deviation of 2.1 Å.

[a] Residues 23–34 of this portion of papain were the initial probe, overlaying residues 120–131 of AVP–pVIc. The r.m.s deviation of the resulting fit was 0.354 Å.

Table 3
Comparison of the distances between residues involved in catalysis in papain (PDB access 1PPN) with comparable residues in AVP–pVIc

Structure	Amino acid residue	Distance (Å)
Papain	Cys^{25} SG–His^{159} ND1	3.80
AVP–pVIc	Cys^{122} SG–His^{54} ND1	3.83
Papain	His^{159} NE2–Asn^{175} OD1	2.55
AVP–pVIc	His^{54} NE2–Glu^{71} OE2	2.63
Papain	Gln^{19} NE2–Cys^{25} N	4.04
AVP–pVIc	Gln^{115} NE2–Cys^{122} N	3.67

may be important in substrate recognition for the P_4 [28] amino acid. Most of the rest of the floor of the active site groove is formed by helix H4, which contains Cys^{122}, the active site nucleophile. Residues 2 to 5 form another wall. Residues 71 to 75 form part of one wall on the far end of the groove. This region contains the third member of the triad, Glu^{71}. It also contains the conserved aromatic amino acids, Phe^{73} and Phe^{75}; in papain in similar positions are Trp^{177} and Trp^{181}. These tryptophan residues of papain are conserved in all similarly folded cysteine proteinases [29,30] and play a major role in the activation of papain. Finally, residues 115 to 120 form part of the other wall of the groove of the AVP–pVIc complex. This wall contains Gln^{115}, the amino acid residue that forms part of the oxyanion hole.

A major difference between the active site grooves of AVP and papain is that in AVP the conserved amino acid residue Trp^{55} emerges from the floor close to Cys^{122}. The predominant amino acid residue at this position in the papain family is Ala^{160}. It is possible that Trp^{55} is a major contributor to the substrate specificity of AVP, and that this may be why its substrate specificity is quite different from that of papain. These differences in the architecture of the two active sites should be exploitable in the design and selection of specific drugs that inhibit enzyme activity.

One reason that AVP represents a new fold is the order along the polypeptide chain of the amino acid residues involved in catalysis. In papain this order is Cys^{25}, His^{159}, Asn^{175}, while in AVP it is His^{54}, Glu^{71}, Cys^{122}. The clear implication from this and other observations is that the AVP–pVIc complex represents a new fold, and AVP–pVIc and papain probably employ the same mechanism of catalysis not because they are related but because they are an example of convergent evolution.

3.4. Active site nucleophiles

In cysteine proteinases, thiol residues with low pK_a values are usually the active site nucleophiles. The mechanism by which a low pK_a is maintained in the cysteine proteinases of the papain family involves the participation of a nearby histidine residue in the active site. The histidine abstracts the proton from the cysteine, resulting in an imidazolium cation and a thiolate anion. In this thiolate–imidazolium ion-pair, the thiolate anion becomes an active site nucleophile [31]. The third member of the catalytic triad in papain is an asparagine residue. The nucleophilic thiolate anion in papain exists over a wide pH range, $pK_{a1} \sim 4.0$ and $pK_{a2} \sim 9.0$. The nucleophilic thiolate anion in cathepsin B has a pK_a of 3.0, that for ficin is 2.5 and for caricain 2.9 [32]. The pH dependency of the alkylation of picornain 2A of rhinovirus type 2 with iodoacetamide shows two reactive thiol forms, the free thiolate ion at high pH and an imidazole assisted thiol group at low pH [33]. Most of the differences in pK_a values can be attributed to differences in the environment of the active site, especially the presence of ionizable groups close to the residues involved in catalysis.

To determine if AVP and AVP–pVIc contain thiols with unusually low pK_a values, AVP and AVP–pVIc were incubated in the presence or absence of iodoacetamide at pH values from pH 3.0 to 8.0, with 0.5 pH unit increments. Under these conditions, only thiolate anions will be alkylated. After the incubations at different pH values, residual enzyme activity was measured, as described in Materials and methods. As a control, similar experiments were performed using papain. The results are expressed as the ratio of the residual enzyme activity observed in enzyme incubated with iodoacetamide to the activity observed in controls not incubated with iodoacetamide (Fig. 5).

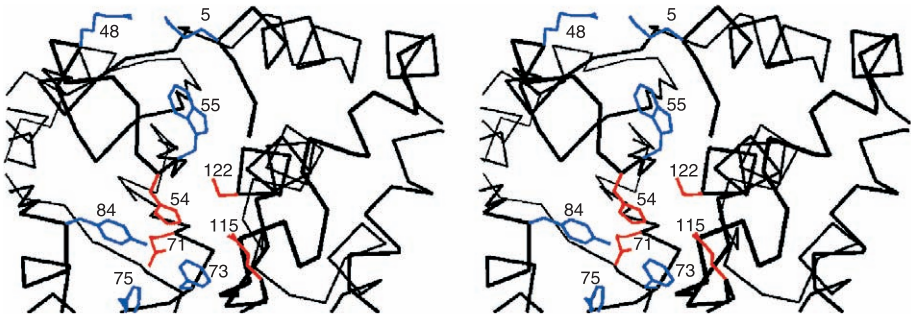

Fig. 4. Cross-eyed stereo view of the active-site region of AVP–pVIc. Conserved residues potentially involved in substrate interactions are colored blue and their sequence position indicated by number. The active site residues are colored red and also numbered.

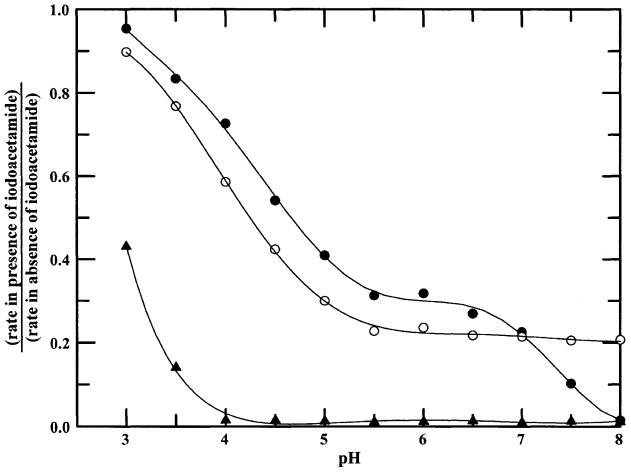

Fig. 5. Carboxamidomethylation as a function of pH. AVP (○), AVP–pVIc complexes (●), or papain (▲) were incubated at the indicated pH in the absence or presence of 2 mM iodoacetamide at ambient temperature for 2 min. Aliquots of the incubations were assayed for enzymatic activity as described in Materials and methods. The results are plotted as the ratio of the observed rate of substrate hydrolysis by the enzyme incubated in the presence of iodoacetamide to that incubated in the absence of iodoacetamide. Assays for AVP or AVP–pVIc complex activity were performed at pH 8.0, while assays for papain activity were performed at pH 5.0.

AVP activity exhibited a complex sensitivity to carboxamidomethylation as a function of pH. One cysteine had a pK_a of about 4.3. AVP–pVIc activity also exhibited a complex sensitivity to carboxamidomethylation as a function of pH. One cysteine had a pK_a of 4.7 and another a pK_a of approximately 7.3. The activity of papain was sensitive to the carboxamidomethylation of a cysteine with a pK_a of 2.9. Under the conditions of the experiment, AVP, AVP–pVIc and papain activities were stable over the pH range used (data not shown). Thus, AVP indeed has a thiol with a low pK_a, much lower than the pK_a of around 8.3 that is expected for a cysteine.

Which cysteine residue in AVP has a pK_a of 4.3 and is essential for enzyme activity? Since five of the eight cysteines of AVP are buried in the interior of the molecule, the most likely candidate to be alkylated was one of the three cysteines found on the surface. Cys^{199} is only found in human adenovirus serotypes 2 and 5; it is located in a region of the molecule distant from either the pVIc binding site or the active site groove. Furthermore, mutation of Cys^{199} to Ser^{199} had no effect on enzyme activity (Brown and Mangel, unpublished). In the AVP–pVIc complex, Cys^{104} forms a disulfide bond with Cys^{10} of pVIc. Thus, the most likely candidate for the low pK_a is the conserved residue Cys^{122}.

The differences in pK_a values observed with iodoacetamide as a function of pH between AVP and AVP–pVIc may be the result of differences in the local environment of the active site. In our model for the activation of AVP by pVIc, we proposed that pVIc acts as a strap that brings the two domains of AVP together. Cys^{122} resides in one domain while His^{54} and Glu^{71} reside in the other. Binding of pVIc to AVP increases the k_{cat} about 350-fold [34]. This indicates the catalytic machinery in AVP is functional, but that upon binding of pVIc, rearrangements occur leading to even greater catalytic efficiency. These rearrangements which do not involve any gross changes in secondary structure [35] resulted in the pK_a of the nucleophile increasing from 4.3 to 4.7. The identity of a second cysteine, with a pK_a of greater than 7.0 whose reaction with iodoacetamide affects AVP–pVIc activity and possibly AVP, is unclear because there are no conserved cysteines, other than Cys^{104} and Cys^{122}, among the 26 AVPs whose sequences have been determined [9]. Alternatively, this could reflect an interaction of iodoacetamide with a histidine residue. The pK_a of a free histidine is 6.0, and the only conserved histidine is His^{54}.

There is now considerable evidence that Cys^{122} is the active site nucleophile of AVP; this amino acid residue is conserved in all adenovirus serotypes. If it is mutated to an alanine residue, enzyme activity is lost [36]. Cys^{122} is located in a deep groove on the surface of AVP–pVIc in close proximity to the conserved His^{54} [14]; Table 3. A 3.83-Å-long hydrogen bond is formed between atoms S of Cys^{122} and Nδ of His^{54}. Not only can His^{54} abstract a proton from Cys^{122}, but Glu^{71} can abstract a proton from His^{54}. A 2.63-Å-long hydrogen bond is formed between atoms Oε of Glu^{71} and Nε of His^{54}. Cys^{122} is at the N terminus of an 11-amino-acid residue α-helix; in that position, the α-helix dipole can interact with the cysteine thiolate anion thereby lowering its pK_a significantly [37]. The active site residues involved in catalysis by papain [38] superimpose with similar residues in AVP [34]. In particular, the Cys–His ion pair in papain superimposes with Cys^{122} and His^{54} of AVP, the distances between the two residues are 3.80 and 3.83 Å, respectively; Table 3. Among adenovirus serotypes, an acidic residue, predominantly aspartate, is always in position 71. A minor difference between AVP and the papain family is that in the papain family the third member of the catalytic triad is always asparagine. Of the three conserved cysteine residues of AVP, Cys^{122} is most likely to have a pK_a of about 4.7, a value similar to the pK_a of the active site nucleophile of papain. The geometry of the catalytic triad of AVP is analogous to that in the serine proteinases [39].

3.5. The interactions between pVIc and AVP

The interactions between AVP and pVIc include main-chain hydrogen bonds accompanied by side-chain hydrogen bonds, ion-pair interactions, van der Waals interactions and a disulfide bond (Fig. 6). There are 24 non-β-strand hydrogen bonds, six β-strand hydrogen bonds, and one covalent bond. A disulfide bond is formed between Cys^{104} of AVP and Cys$^{10'}$ of pVIc. Between pVIc and AVP with its solvent shell, there are 30 hydrogen bonds, four of which are ion-pairs, as detailed in Table 4.

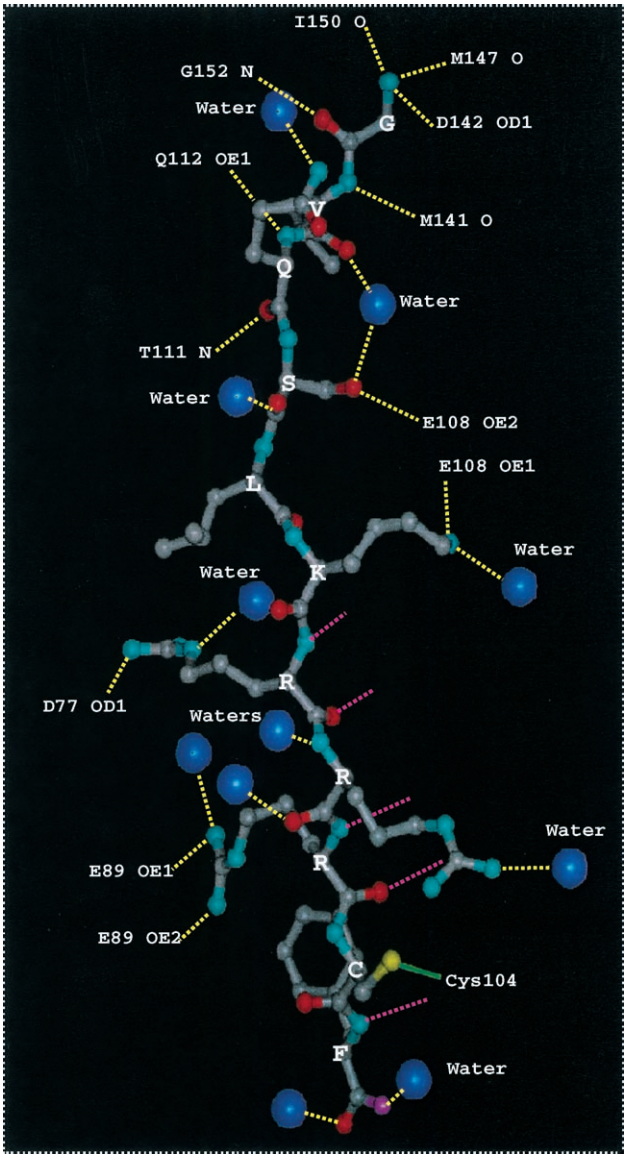

Fig. 6. Binding interactions between pVIc and AVP. There are 24 non-β-strand hydrogen bonds, one covalent bond and six β-strand hydrogen bonds between pVIc and AVP. pVIc is shown in ball and stick form with carbon atoms in gray, oxygen atoms in red, and nitrogen atoms in cyan. Waters are represented by blue spheres, and hydrogen bonds to the indicated AVP residues are dashed yellow lines. β-strand hydrogen bonds are dashed magenta lines. The disulfide bond is shown as a solid green line. This view does not show the first β-strand hydrogen bond, located directly beneath the Leu$^{5'}$ oxygen.

Surprisingly, pVIc, which exerts powerful control on the rate of catalysis, binds quite far from the active-site residues involved in catalysis. pVIc binds on the front-bottom edge of AVP, away from the active-site groove on the top (Fig. 2A). The disulfide bond between AVP Cys^{104} and pVIc Cys$^{10'}$ is 32 Å from Cys^{122}, the active site nucleophile. The residue of pVIc that is closest to the active site is the side-chain of Val$^{2'}$, which is about 14 Å from Cys^{122}. How could the binding of pVIc to AVP alter the rate of substrate hydrolysis by the enzyme? The four N-terminal residues of pVIc interact, via hydrogen bonds and van der Waal's attractions, with residues located in the helical domain of the AVP molecule, interacting at a distance with one of the bridges between the domains. The next five residues of pVIc cross between the two domains, into the β-sheet domain, forming the sixth strand of the central β-sheet. The penultimate amino acid, Cys$^{10'}$, of pVIc is covalently linked via a disulfide bond to Cys^{104} of AVP, and the C-terminal Phe$^{11'}$ side chain is locked into a hydrophobic pocket formed in part because of side chain interactions of pVIc with AVP. Taken together, these interactions suggest that the pVIc is acting as a strap to facilitate the alignment of the two domains of AVP, allowing for more efficient catalysis.

3.6. Conserved amino acids provide some insights on the fold

The nucleic acid sequences of 24 AVP genes have been published [9]. Inspection of the alignment of the translated open reading frames, using FASTA, reveals that amino acids at 33 positions are conserved. Among those are the amino acid residues proposed to constitute the catalytic triad, His^{54}, Cys^{122}, and the carboxylate at position 71. At

Table 4
Hydrogen bonds between pVIc and AVP and its solvent shell

Donor	Acceptor	Distance (Å)
AVP152 N	pVIc 1′ O	2.92
pVIc 1′ N	AVP 142 OD1	2.86
pVIc 1′ N	AVP 147 O	2.98
pVIc 1′ N	AVP 150 O	2.71
pVIc 2′ N	AVP 141 O	2.99
AVP 1001 OH_2	pVIc 2′ O	2.84
pVIc 3′ N	AVP 112 OE1	2.99
pVIc 3′ NE2	AVP 1012 OH_2	2.99
AVP 111 N	pVIc 3′ O	2.92
pVIc 4′ OG	AVP 108 OE2	2.66
pVIc 4′ OG	AVP 1001 OH_2	2.85
AVP 1220 OH_2	pVIc 4′ O	2.57
pVIc 5′ N	AVP 109 O	2.94
pVIc 6′ NZ	AVP 108 OE1	2.53
pVIc 6′ NZ	AVP 1151 OH_2	2.92
pVIc 7′ N	AVP 107 O	2.76
pVIc 7′ NE	AVP 1105 OH_2	2.91
pVIc 7′ NH2	AVP 77 OD1	2.87
AVP 107 N	pVIc 7′ O	2.84
pVIc 8′ N	AVP 1095 OH_2	2.79
pVIc 8′ NH2	AVP 1047 OH_2	2.90
AVP 1164 OH_2	pVIc 8′ O	2.95
pVIc 9′ N	AVP 105 O	2.86
pVIc 9′ NH1	AVP 89 OE2	2.76
pVIc 9′ NH2	AVP 89 OE1	2.89
pVIc 9′ NH2	AVP 1131 OH_2	2.89
AVP 105 N	pVIc 9′ O	2.79
pVIc 11′ N	AVP 103 O	2.78
AVP 1168 OH_2	pVIc 11′ O	2.73
AVP 1072 OH_2	pVIc 11′ OXT	2.62

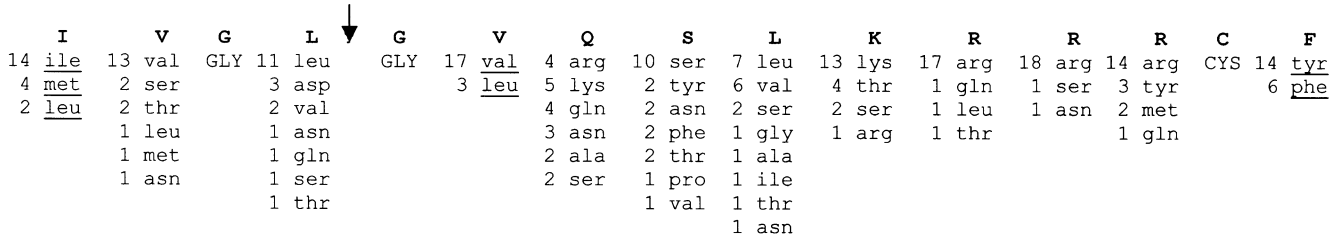

Fig. 7. Conserved amino acids at the C terminus of virion precursor protein pVI. Sequences of pVI, from the AVP consensus cleavage sequence to the C terminus, are shown from 20 strains of adenovirus. Sequence identity is indicated by capital letters; sequence homology is indicated by underlining. Cleavage of pVI at the arrow releases pVIc.

position 71, there is an aspartate for all adenovirus serotypes, except 2 and 5, where there is a glutamate.

Several conserved amino acid residues aid in the forming of the active site groove and, in doing so, the alignment of His^{54}. The most notable of these interactions involves the residue Thr^{45}. Its nitrogen atom forms a hydrogen bond to the His^{54} oxygen atom. The side chain of Thr^{45} forms part of a hydrophobic pocket that includes parts of the His^{54} imidazole ring and the side chains of Met^{56}, Tyr^{84}, and Phe^{86}. Met^{56} is present in serotypes 2 and 5; in 15 other serotypes it is a leucine, in six others it is an isoleucine. Tyr^{84} is conserved among adenovirus serotypes. The phenylalanine at position 86 is replaced by valine in one serotype and tyrosine in two others. All these substitutions would maintain a hydrophobic pocket and orient the side chain of His^{54}.

Several residues, either conserved or replaced by homologous residues, are concerned with the formation of the active site groove and with substrate specificity. Among these are Glu^{5}, Arg^{48}, and Trp^{55}. Glu^{5} and Arg^{48} form an ion pair that closes one end of the groove. With Trp^{55}, they form a pocket in the floor of the groove. This pocket may be involved in substrate specificity, in particular the binding of the P_4 amino acid residue in a substrate. Also, Gly^{123} is only 3.6 Å from the side chain of Trp^{55}; these residues may also be involved in the P_2 binding site.

Most of the other conserved amino acid residues are involved in structural roles. All the conserved aromatic residues, phenylalanines at positions 21, 29, 32, 70, 133, and 185 and tyrosines at positions 84 and 175, interact between secondary structure elements. The aromatic side chains of phenylalanines 29 and 32 stack against the central β-sheet between strands S2 and S3. Phenylalanine 133 stacks against the central β-sheet from the opposite side, also at strands S2 and S3. The aromatic rings of Tyr^{175} and Phe^{185}, along with Thr^{196}, stack between helices H5 and H6.

pVIc has two conserved amino acid residues—$Gly^{1'}$ and $Cys^{10'}$ [40] (Fig. 7). At three other positions, no more than two different amino acids are allowed. At pVIc2′, there is predominantly Val with three Leu. At pVIc8′, all are Arg except for one Asn and one Ser. At pVIc11′, 14 are Tyr and six are Phe. The three most substituted amino acid residues are at positions 3′, 4′, and 5′.

3.7. Proteins homologous to AVP

Until recently, AVP appeared to be unique. The sequence of the gene was unique; the fold of the proteinase was unique. Then, a tandem pair of paralogous genes was identified in *Chlamydia* that were predicted to have a fold like that of AVP [41]. A proteinase required for cell-cycle progression in yeast, ubiquitin-like protease 1 (ulp1), was shown to be homologous to AVP [12]. The enzyme cleaves SUMO–protein fusions and/or isopeptide-bound-linked SUMO–protein conjugates. A virulence factor in *Yersinia pestis*, YopJ, has structural homology to AVP [42]. The primary sequences of the enzymes from *Chlamydia* and *Yersinia* contain the same catalytic triad and oxyanion hole amino acids in the same relative positions along the polypeptide chain as they are in AVP. A protease related to AVP has been found in *Saccharomyces cerevisiae* that is involved in the regulation of chromosome condensation [43].

Acknowledgements

We thank John Skinner and Salvatore Sclafani of beamline X12-C at the National Synchrotron Light Source for their valuable assistance. We also thank Mary Lynn Baniecki and Diana Toledo for helpful discussions. Research was supported by the Office of Biological and Environmental Research of the US Department of Energy under Prime Contract no. DE-AC0298CH10866 with Brookhaven National Laboratory, and by National Institutes of Health Grant AI41599. A.D. was supported by the Department of Energy's Office of Science Education and Technical Information, as a Science and Engineering Research Semester Program participant.

References

[1] J. Weber, J. Virol. 17 (1976) 462–471.
[2] M.T. Brown, W.J. McGrath, D.L. Toledo, W.F. Mangel, FEBS Lett. 388 (1996) 233–237.
[3] C.W. Anderson, P.R. Baum, R.F. Gesteland, J. Virol. 12 (1973) 241–252.
[4] L. Yeh-Kai, G. Akusjarvi, P. Alestrom, U. Pettersson, M. Tremblay, J. Weber, J. Mol. Biol. 167 (1983) 217–222.

[5] W.F. Mangel, W.J. McGrath, D.L. Toledo, C.W. Anderson, Nature 361 (1993) 274–275.
[6] C.W. Anderson, Protein Expr. Purif. 4 (1993) 8–15.
[7] A. Webster, G. Kemp, J. Gen. Virol. 74 (1993) 1415–1420.
[8] W.J. McGrath, M.L. Baniecki, C. Li, S.M. McWhirter, M.T. Brown, D.L. Toledo, W.F. Mangel, Biochemistry 40 (2001) 13237–13245.
[9] C. Barbezange, M. Benko, A. Dan, B. Harrach, Virus Res. 6 (2000) 79–85.
[10] G. Akusjarvi, H.J. Persson, J. Virol. 38 (1981) 469–482.
[11] F. Cai, J.M. Weber, Virology 196 (1993) 358–362.
[12] S.-J. Li, M. Hochstrasser, Nature 398 (1999) 246–251.
[13] W.J. McGrath, A.P. Abola, D.L. Toledo, M.T. Brown, W.F. Mangel, Virology 217 (1996) 131–138.
[14] J. Ding, W.J. McGrath, R.M. Sweet, W.F. Mangel, EMBO J. 15 (1996) 1778–1783.
[15] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalovk, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[16] D. Leung, G. Abbenante, D.P. Fairlie, J. Med. Chem. 43 (2000) 305–341.
[17] A. Velazquez-Campy, M.J. Todd, E. Freire, Biochemistry 39 (2000) 2201–2207.
[18] W.F. Mangel, D.L. Toledo, M.T. Brown, J.H. Martin, W.J. McGrath, J. Biol. Chem. 271 (1996) 536–543.
[19] S.G. Gill, P.H. von Hippel, Anal. Biochem. 182 (1989) 319–326.
[20] P.S. Riddles, R.L. Blakeley, B. Zerner, Anal. Biochem. 94 (1979) 75–81.
[21] W.J. McGrath, J. Ding, R.M. Sweet, W.F. Mangel, J. Struct. Biol. 117 (1996) 77–79.
[22] Z. Otwinowski, W. Minor, Methods Enzymol. 276 (1997) 307–326.
[23] A.T. Brünger, XPLOR Manual Version, Yale Univ. Press, New Haven, CT, 1992.
[24] G.M. Sheldrick, Acta Crystallogr., A 46 (1990) 467–473.
[25] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J. Appl. Crystallogr. 26 (1993) 283–291.
[26] G.J. Kleywegt, A.T. Brunger, Structure 4 (1996) 897–904.
[27] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Acta Crystallogr., A 47 (1991) 110–119.
[28] I. Schechter, A. Berger, Biochem. Biophys. Res. Commun. 27 (1967) 157–162.
[29] I.G. Kamphuis, J. Drenth, E.N. Baker, J. Mol. Biol. 182 (1985) 317–329.
[30] A. Ritonja, T. Popvic, M. Kotnik, W. Machleidt, V. Turk, FEBS Lett. 228 (1988) 341–345.
[31] L. Polgar, FEBS Lett. 47 (1974) 15–18.
[32] S. Pinitglang, A.B. Watts, M. Patel, J. Reid, D.M.A. Noble, S. Gul, A. Bokth, A. Naeem, H. Patel, E.W. Thomas, S.K. Sreedharan, C. Verma, K. Brocklehurst, Biochemistry 36 (1997) 9968–9982.
[33] Z. Sarkany, T. Skern, L. Polgar, FEBS Lett. 481 (2000) 289–292.
[34] W.F. Mangel, D.L. Toledo, J. Ding, R.M. Sweet, W.J. McGrath, Trends Biochem. Sci. 22 (1997) 393–398.
[35] M.L. Baniecki, W.J. McGrath, S.M. McWhirter, C. Li, D.L. Toledo, P. Pellicena, D.L. Barnard, K.S. Thorn, W.F. Mangel, Biochemistry 40 (2001) 12349–12356.
[36] A.W. Grierson, R. Nicholson, P. Talbot, A. Webster, G.D. Kemp, J. Gen. Virol. 75 (1994) 2761–2764.
[37] T. Kortemme, T.E. Creighton, J. Mol. Biol. 253 (1995) 799–812.
[38] J. Drenth, J.N. Jansonius, R. Koekoek, B.G. Wolthers, in: P.D. Boyer (Ed.), The Enzymes, vol. 3, Academic Press, New York, 1971, pp. 484–500.
[39] D.M. Blow, J.J. Birktoft, B.S. Hartley, Nature 221 (1969) 337–340.
[40] A. Ruzindana-Umunyana, S. Sircar, J.M. Weber, Virology 270 (2000) 173–179.
[41] R.S. Stephens, S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R.L. Tatusov, Q. Zhao, E.V. Koonin, R.W. Davis, Science 282 (1998) 754–759.
[42] K. Orth, Z. Xu, M.B. Mudgett, Z.Q. Bao, L.E. Palmer, J.B. Bliska, W.F. Mangel, B. Staskawicz, J.E. Dixon, Science 290 (2000) 1594–1597.
[43] A.V. Strunnikov, L. Aravind, E.V. Koonin, Genetics 158 (2001) 95–107.